

# 基于Django的陶瓷商品评论情感分析系统设计

周立斌, 谢志洋, 王宇凯

景德镇陶瓷大学信息工程学院, 江西 景德镇

收稿日期: 2024年4月25日; 录用日期: 2024年5月27日; 发布日期: 2024年6月24日

## 摘要

随着互联网的兴起, 各大电商平台上的商品评价对消费者的购买决策产生了重大影响, 特别是在购买文化和艺术价值较高的陶瓷商品时。为了帮助消费者和企业更好地利用和分析这些评论信息, 本文设计并实现了一个基于Django框架的陶瓷商品评论情感分析系统, 该系统通过数据挖掘技术和机器学习算法对消费者评论进行模型的训练和情感分析。系统后端使用Python进行开发, 前端则利用Bootstrap和Pychart等技术, 实现了对分析结果的直观可视化展示, 帮助消费者更全面地了解所要购买的商品。

## 关键词

陶瓷, 情感分析, Django, 模型训练, 可视化展示

# Design of a Sentiment Analysis System for Ceramic Product Reviews Based on Django

Libin Zhou, Zhiyang Xie, Yukai Wang

School of Information Engineering, Jingdezhen Ceramics University, Jingdezhen Jiangxi

Received: Apr. 25<sup>th</sup>, 2024; accepted: May 27<sup>th</sup>, 2024; published: Jun. 24<sup>th</sup>, 2024

## Abstract

With the rise of the Internet, product reviews on various e-commerce platforms have significantly influenced consumer purchasing decisions, especially when buying ceramics with high cultural and artistic value. To assist consumers and businesses in better utilizing and analyzing these reviews, this paper designs and implements a sentiment analysis system for ceramic product reviews based on the Django framework. This system employs data mining techniques and machine learning algorithms to train models on consumer reviews and conduct sentiment analysis. The backend of the system is developed using Python, while the frontend utilizes technologies like Bootstrap and Pyechart to visually display the analysis results, helping consumers gain a comprehensive understanding of the products they intend to purchase.

## Keywords

Ceramics, Sentiment Analysis, Django, Model Training, Visualization Display

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

互联网技术的发展推动了电子商务的繁荣，使得消费者可以在家中轻松购买全球商品。在众多在线商品中，陶瓷制品以其独特的文化和艺术价值成为热门选择。然而，由于线上无法直观体验商品实物，在这种情境下，商品评论的作用变得尤为重要。它们不仅帮助消费者了解商品的实际表现，还影响着消费者的购买决策和商家的信誉。然而，由于评论数量庞大且内容复杂，手动分析这些评论既耗时又低效。因此，本研究提出了构建一个基于 Django 的陶瓷商品评论情感分析系统，该系统利用先进的数据挖掘技术和机器学习算法，自动化地处理和分析用户评论，进而提供情感倾向分析和关键信息提取。此情感分析系统的开发不仅可以为消费者提供更为直观、全面的商品信息，帮助他们做出更明智的购买选择，也为电商平台和陶瓷制造商提供了宝贵的消费者反馈，支持他们改进产品和服务，优化用户体验。随着技术的进步和市场的不断扩展，此类情感分析工具的重要性将持续增长，它们将在电子商务领域扮演越来越关键的角色。通过本系统的应用，期望能够为电商行业的持续创新和发展提供支持，同时为消费者和商家创造更多价值。

## 2. 系统整体框架设计

在设计基于 Django 的陶瓷商品评论情感分析系统时，我们采用模块化的策略，细分系统为数据预处理、模型训练和可视化情感分析三个关键子模块。各模块间关系如图 1 所示。初始阶段，由于电商平台中陶瓷商品的评论往往包含大量脏数据，例如拼写错误、语法问题及各类特殊符号，这些都可能对后续分析的准确性造成负面影响。为了解决这一问题，数据预处理模块首先进行了综合的数据清理，包括纠正拼写错误、规范化语法、去除无关符号以及停用词的剔除。此外，该模块还实现了对评论文本的分词和关键词提取，为深度自然语言处理奠定基础。

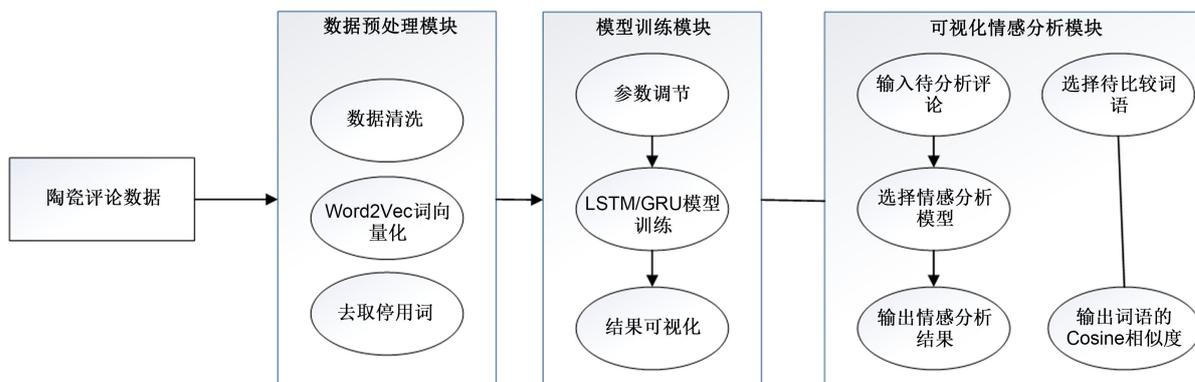


Figure 1. Relationship between modules

图 1. 模块间关系

进一步地，我们采用了 Word2Vec 技术对处理后的文本数据进行向量化。Word2Vec 的应用不仅优化了数据的表示方式，还为机器学习模型的输入提供了高质量的特征集。基于这些特征集，模型训练模块随后开展了两种主流的深度学习模型的训练工作：门控循环单元(GRU)和长短期记忆网络(LSTM)。这两种模型都属于循环神经网络(RNN)的变体，特别适合处理时间序列数据如文本评论。它们通过学习评论中的语义和情绪变化，能够有效地从文本中识别情感倾向[1]。

在模型训练完成后，系统借助训练精细化的情感分析模型和 Word2Vec 模型，通过一个直观的可视化情感分析模块为用户展示模型训练结果。用户能够针对特定的评论内容执行即时分析，也可以与预先训练的 Word2Vec 模型实现可视化交互，简化了复杂数据的解读。

### 3. 系统关键功能模块

#### 3.1. 数据预处理模块

在数据预处理模块中，我们关注的首要任务是清理评论文本中的非中文字符，为此，我们采用了 Python 中的正则表达式，利用 `re.sub()` 函数定义了过滤规则，有效地将这些干扰因素从数据中清除，确保仅留下有用的中文文本供后续分析使用。紧接着，jieba 分词工具的引入进一步提高了文本处理的准确度，它使我们能够迅速而精准地将连续的文本字符串切成有意义的单词或短语。在应用 jieba 的 `lcut()` 方法进行分词的同时，我们通过加载哈工大停用词库，以便从文本中剔除“这”、“顺着”、“的”等词汇，这些词通常对情感分析贡献不大。通过这些步骤，数据预处理模块可以显著提高后续模型训练和情感分析的质量和效率。

#### 3.2. 模型训练模块

在模型训练模块的构建中，本研究训练了 LSTM 和 GRU 这两种循环神经网络结构，以准确处理和预测文本数据中的情感标签。LSTM 模型以其在捕捉长期依赖信息方面的出色能力，为理解文本的上下文提供了强大支持。模型通过嵌入层、LSTM 层和全连接层的结构组合，确保了对情感的精确分类。训练中采用了 Adam 优化器和交叉熵损失函数，强化了模型在实际应用中的性能和泛化能力。GRU 模型以更简洁的结构和减少的参数量，展现了在某些情境下比 LSTM 更高效的训练和收敛速度，优化了情感分析结果[2]。两模型的训练和评估过程都经过了严格的数据分割和周期性评价，以保证模型的准确率和泛化性。最终，训练好的模型结构和权重被妥善保存，为网站可视化情感分析模块提供了模型基础。

#### 3.3. 可视化情感分析模块

本研究中的可视化情感分析模块结合了自然语言处理与前端技术，为用户提供了一个直观且交互性强的平台，如图 2 所示，用于即时评估陶瓷商品评论的情感倾向。模块背后的核心是基于 LSTM 和 GRU 算法的神经网络模型，这些模型能够识别文本中的情感标签并将其分类为“好评”、“中评”或“差评”。同时，借助预训练的 Word2Vec 模型，系统还能展示不同词汇间的相似性，进一步丰富了用户的分析维度。用户在文本框中输入评论后，可以选择使用 LSTM 或 GRU 模型来分析文本情感。系统后端，利用 Python 进行了一系列的数据处理，例如使用正则表达式清洗数据和 jieba 进行分词，确保输入评论数据的质量后。通过 Django 框架配合 Ajax 技术，实现了与前端的无缝连接，用户的交互请求可以实时传递给后端，后端处理完成后，情感分析模型结果会立即反馈到前端页面上。

### 4. 平台开发技术实现

在本研究构建的陶瓷商品评论智能化分析系统中，综合运用了多种技术来打造一个全栈式解决方案。系统的核心框架选用了 Django，这是一个高度成熟且功能丰富的 Python Web 框架，它为我们的后端逻辑

提供了稳健的支持，特别适合快速开发有复杂数据处理需求的 Web 应用。前端开发方面，我们利用了 HTML5、CSS3 与 JavaScript，以确保用户界面的现代化和互动性，结合 Bootstrap 框架实现了响应式设计。为实现前后端的高效数据交互，Ajax 技术被广泛应用于用户的实时请求处理。系统的情感分析功能建立在 NLTK、jieba 分词和 Gensim 等自然语言处理工具上[3]，而 Word2Vec 模型的引入则进一步提升了词汇处理的精度。机器学习模型的构建和训练选用了 Keras 和 TensorFlow，这两个库在深度学习社区中有着广泛的应用和支持。数据持久化方面，MySQL 数据库被选中以存储用户数据和评论内容，其健壮的性能保证了数据操作的效率和安全性。通过如上的技术组合，我们构建了一个陶瓷商品评论智能化分析平台。



Figure 2. Visual use of sentiment analysis model  
图 2. 情感分析模型可视化使用

#### 4.1. Django 框架搭建

在搭建本智能分析系统时，我们选择了 Django 框架作为 Web 开发的核心。首先 Django 的模型 - 视图 - 模板(MVT)架构为我们的应用提供了良好的结构，它使得业务逻辑、数据处理与用户界面得以高效分离[4]，极大地提高了代码的整洁性与可维护性。特别是在处理复杂且庞大的陶瓷商品评论数据时，Django 的对象关系映射(ORM)系统显得尤为重要，它通过高级的抽象简化了数据库的操作，让数据库的创建、操作和维护变得既直观又灵活。借助 Django 强大的 ORM 系统，我们无需深入 SQL 语言的复杂性，就能够完成复杂查询和数据管理任务。这不仅加快了开发速度，而且在后期进行数据结构调整时也为我们带来了极大的便利，使数据库的迁移和修改变得更加简单和安全。其次，Django 框架内置了一系列安全措施，包括但不限于用户认证、密码管理、会话安全以及跨站请求伪造(CSRF)保护[5]。这些功能为平台提供了坚固的防御，使其能够抵御包括 SQL 注入、身份盗用和其他网络攻击在内的常见安全威胁。这些技术决策初步支撑起了一个陶瓷商品评论智能化分析系统

#### 4.2. 前端界面搭建

平台采用了 HTML5、CSS3 和 JavaScript 技术，结合 Bootstrap 框架实现响应式设计，旨在为用户提供一个既现代又交互性强的使用平台。在 2.2 模型训练模块用户可以根据自己的需要自定义情感分析模型的训练。界面中集成了模型训练的各种参数设置选项，如选择不同的神经网络结构(LSTM 或 GRU)、

设置迭代次数、批次处理的样本数量等。为了提升系统的互动性和用户的操作便捷性，我们在前端实现了多种 Web 交互技术，例如通过 Ajax 技术实现页面的异步更新，用户的操作请求可以不刷新页面即时传递给后端，后端处理完成后立即将结果反馈到前端显示，大大提升了系统的响应速度[6]。

### 4.3. 情感分析模型搭建

在深度学习模型搭建这一部分，我们关注的是构建适用于文本分类任务的 LSTM 和 GRU 模型。这两种模型是常用的循环神经网络(RNN)架构，能够有效地捕捉文本数据中的长期依赖关系，适用于像陶瓷评论这样具有长短不一的句子结构和变长的序列。我们采用 Python 的 Keras 和 TensorFlow 库来构建和训练模型，这些库提供了丰富的工具和接口，极大地简化了深度学习模型的开发流程[7]。在模型结构上，我们选择了顺序模型(Sequential Model)，由于其结构简单、易于管理，非常适合处理多种深度学习任务。在 LSTM 模型中，首先通过一个嵌入层(Embedding Layer)将输入的词汇转换为密集的向量表示，然后通过两个包含 100 个单元的 LSTM 层来处理这些向量，这样做可以利用 LSTM 的记忆单元结构有效捕捉文本中的时间序列依赖关系。在模型的末端，我们加入了一个全连接层[7]，并使用 softmax 激活函数来执行多类别分类任务。GRU 模型也采用了类似的结构，但它简化了内部结构，减少了参数数量，使得训练过程更为高效，尽管性能略有差异。在训练和评估阶段，我们将数据集分为训练集和测试集，使用交叉熵损失函数和 Adam 优化器，细致监控模型在测试集上的性能，如准确率、召回率和 F1 分数等指标，结果如表 1 所示，确保模型具备良好的泛化能力。最后保存训练好的模型，以便后续可视化情感分析模块的预测和应用。此外，我们利用 Plotly 库绘制训练和验证过程中的准确率和损失曲线，以直观展示模型性能的变化。通过构建混淆矩阵如图 3 所示，进一步分析模型的分类效果，并计算精确度、召回率和 F1 分数，以全面评估模型性能。这些步骤不仅帮助我们调整和优化模型参数，也确保了模型在实际应用中的可靠性和效果。

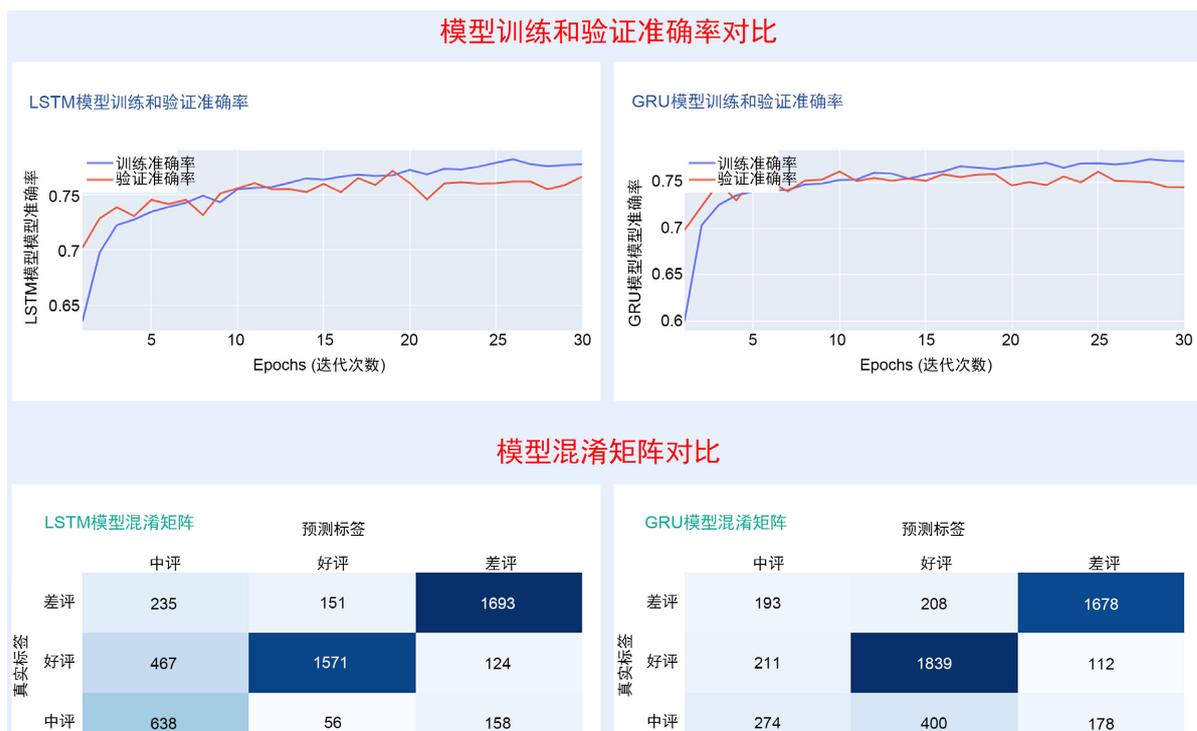


Figure 3. Confusion matrix and accuracy curve

图 3. 混淆矩阵和准确率曲线

**Table 1.** Values of model indexes**表 1.** 模型各项指标值

模型	准确率	召回率	F1 分数
LSTM	0.8046	0.7661	0.7768
GRU	0.7346	0.7443	0.7371

## 5. 结论

本研究成功设计并实现了一个基于 Django 的陶瓷商品评论智能化分析系统，旨在提供一个高效、实用的平台，以支持消费者和企业更好地理解 and 利用在线评论信息。通过结合自然语言处理技术和深度学习模型[8]，特别是 LSTM 和 GRU，本系统能够准确地分析和分类评论中的情感倾向，从而帮助用户做出更为明智的购买决策和市场策略。系统的开发集成了数据预处理、模型训练、情感分析和结果可视化等关键模块。通过自动化的数据清洗和预处理，系统能够有效地处理和转化原始评论数据，为后续的深度学习模型训练提供了高质量的输入。利用现代的机器学习框架，我们训练了精准的情感分析模型，这些模型不仅展示了优异的分类性能，还通过直观的图形和报表，使得分析结果易于理解和操作。在未来的工作中，我们计划进一步扩展系统的功能，包括增加更多的语言支持和复杂的数据分析工具，以适应更广泛的市场需求和用户偏好。此外，考虑到 AI 和机器学习技术的快速发展，我们也将探索更先进的模型和算法，以持续提高系统的性能和准确性。

## 基金项目

景德镇陶瓷大学大学生创新创业项目资助(编号：202310408018)。

## 参考文献

- [1] 王剑辉, 闫芳序. 基于 LSTM + Word2vec 的微博评论情感分析[J/OL]. 沈阳师范大学学报(自然科学版), 2024(2): 138-144. <http://kns.cnki.net/kcms/detail/21.1534.N.20240424.1638.012.html>, 2024-04-25.
- [2] 王泽宇, 张志清. LSTM 和 GRU 模型对湖北省物流需求预测性能比较研究[J]. 物流工程与管理, 2024, 46(4): 10-14.
- [3] 杨凯华, 刘勇. 基于 Bootstrap 框架的响应式网页设计与实现[J]. 大众标准化, 2024(4): 43-45.
- [4] 鲁程逸, 代子正, 王艳红. 基于情感分析的电商平台评论应用研究[J]. 科技创业月刊, 2023, 36(10): 150-154.
- [5] 郭鹤楠. 基于 Django 和 Python 技术的网站设计与实现[J]. 数字通信世界, 2023(6): 60-62.
- [6] 赵颖. 基于 Ajax 的计算机 Web 请求异步处理技术[J]. 数字技术与应用, 2023, 41(6): 154-157. <https://doi.org/10.19695/j.cnki.cn12-1369.2023.06.48>
- [7] 谢思雅, 施一萍, 胡佳玲, 等. 基于 Django 的文本情感分类系统设计与实现[J]. 传感器与微系统, 2021, 40(11): 97-99. [https://doi.org/10.13873/J.1000-9787\(2021\)11-0097-03](https://doi.org/10.13873/J.1000-9787(2021)11-0097-03)
- [8] 蹇常林. ORM 在 Django 操作数据库中的应用[J]. 技术与市场, 2020, 27(1): 56-57.